02/18/94 12:02 ☎202 404 7478 AFOSR/NL ··· ···

AD-A278 412

‖‖‖‖‖‖‖‖‖‖‖

DOCUMENTATION PAGE *p 2* 80014

Form Approved
OMB No. 0704-0188

②

[fine print block — illegible]

| | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
| | 3-1-94 | Final Report 15 Jan 93- 14 Jan 94 |

4. TITLE AND SUBTITLE

Visual Perception: Computation and Psychophysics

6. AUTHOR(S)

David C. Knill w/ Whitman Richards (MIT)

5. FUNDING NUMBERS

C/G F49620-93-1-0124
61102F
2313
AS

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Minnesota, ORTTA
1100 Washington Ave., Suite 201
Minneapolis, MN 55415

8. PERFORMING ORGANIZATION REPORT NUMBER

DTIC

AFOSR-TR- 94 0219

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NL
110 DUNCAN AVE SUITE B115
BOLLING AFB DC 20332-0001

Dr. John F. Tangney

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC ELECTE SB D
APR 20 1994

11. SUPPLEMENTARY NOTES

Conference grant to support a Workshop held January, 1993.
Final report consists of talk abstracts and outline of resulting book

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The workshop brought together researchers in computational vision and
psychophysics to discuss ways of conceptualizing and modeling problems
in visual perception. Such a conceptualization requires common frameworks
for formulating problems in perception. Workshop participants considered
what formal tools and structures these frameworks should provide in
order to be most useful for the study of human vision. Several recently
proposed frameworks based on the formulation of Bayesian, probablistic,
inference served as the focal point for evaluation and discussion.

14. SUBJECT TERMS

Vision, perception, computation, psychophysics, Bayes,
probabilistic inference

15. NUMBER OF PAGES

15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std. 239-18
298-102

94-11915

DTIC QUALITY INSPECTED 3

94 4 19 045

15 MAR 1994

FINANCIAL STATUS REPORT
(Short Form)
(Follow instructions on the back.)

1. Federal Agency and Organizational Element to Which Report is Submitted U.S. AIR FORCE OFFICE OF SCIENTIFIC RESEARCH	2. Federal Grant or Other ID Number Assigned By Federal Agency F49620-93-1-0124	OMB approval No. 0348-0039	Page 1	of 1

3. Recipient organization (Name and complete address, including ZIP code.)

University of Minnesota, ORTTA
1100 Washington Avenue, Suite 201
Minneapolis, MN 55415

4. Employer Identification Number 1416007513-	5. Recipient account number or ID number 0493-1223 1691-485-9002	6. Final Report |X| Yes | | No	7. Basis |X| Cash | | Accrual

8. Funding/Grant Period From: (Month, Day, Year) 01/01/93	To: (Month, Day, Year) 12/31/93	9. Period covered by this report From: (Month, Day, Year) 01/01/93	To: (Month, Day, Year) 12/31/93

10. Transactions:	I Previously Reported	II This Period	III Cumulative
a. Total Outlays	0.00	9,982.00	9,982.00
b. Recipient share of outlays	0.00	0.00	0.00
c. Federal share of outlays	0.00	9,982.00	9,982.00
d. Total unliquidated obligations			0.00
e. Recipient share of unliquidated obligations			0.00
f. Federal share of unliquidated obligations			0.00
g. Total Federal share (Sum of lines c and f)			9,982.00
h. Total Federal funds authorized for this funding period			9,982.00
i. Unobligated balance of Federal funds (Line h − line g)			0.00

11. INDIRECT EXPENSE	a. Type of rate (Place an 'X' in appropriate box) |X| Provisional | | Predetermined | | Final | | Fixed				
	b. Rate 10%TDC	c. Base 9,074.55	d. Total Amount 907.45	e. Federal Share $907.45	

12. Remarks: Attach any explanations deemed necessary or information required by Federal sponsoring agency in compliance with governing legislation.

13. Certification: I certify to the best of my knowledge and belief that this report is correct and complete and that all outlays and unliquidated obligations are for the purposes set forth in the award documents.

Typed or printed name and title Sandra Kenyon, Grants & Contracts Accounting Manager Office of Research and Technology Transfer	Telephone (Area code, number and extension) (612) 624-6026
Signature of authorized certifying official *Sandra Kenyon*	Date report submitted 02/23/94

Previous Editions not Usable

GB

Workshop on Visual Perception: Computation and Psychophysics†

Jan. 14-17, 1993

Program

Workshop on Visual Perception: Computation and Psychophysics

Chatham Bars Inn
Chatham, MA
(on Cape Cod)

Jan. 14-17

Organized by

David C. Knill, Dept. of Psychology, Univ. of Minnesota, Minneapolis, MN 55455 (612) 626-7518   knill@eye.psych.umn.edu

Whitman Richards, Dept. of Brain and Cognitive Sciences, Mass. Institute of Technology, Cambridge, Mass. 02139 (617) 253-5776   whit@ai.mit.edu

Alan Yuille, Division of Applied Sciences, Harvard Univ., Cambridge, Mass. 02138 (617) 495-9526   yuille@gramian.harvard.edu

Heinrich Bulthoff, Dept. of Cognitive and Linguistic Sciences, Brown Univ., Box 1987, Providence, RI 02912 (401) 863-3383   hhb@holmes.cog.brown.edu

Abstract:

The Workshop brings together researchers in computational vision and psychophysics to discuss ways of conceptualizing and modeling problems in visual perception. Such a conceptualization requires common frameworks for formulating problems in perception. Workshop participants will consider what formal tools and structures these frameworks should provide in order to be most useful for the study of human vision. Several recently proposed frameworks based on the formalism of Bayesian, probabilistic inference will serve as the focal point for evaluation and discussion.

Motivation:

In the decade since Marr's seminal work *Vision: A computational investigation into the human representation and processing of visual information* , advances in computational vision and psychophysics have led to many changes in the way we conceptualize and study problems in visual perception. Despite these advances, most fruitful interactions between computational and psychophysical work have been, with some notable exceptions, limited to the study of low-level visual mechanisms. Much less common are studies which integrate computational and psychophysical approaches to problems in higher-level visual processing such as object recognition, shape perception, cue integration and cooperative estimation of scene properties. Given the progress made on these problems in computational vision and psychophysics over the last decade, the time is ripe to bring researchers in these fields together to discuss ways of conceptualizing and modeling problems in vision which support the integration of knowledge from computational and psychophysical studies. A prerequisite for such integration is the existence of common frameworks for formulating problems in vision.

Recently, a number of groups of researchers have developed formal frameworks for vision built on the principles of Bayesian, probabilistic reasoning. These were developed with an eye toward providing a common language and set of formal tools for specifying both computational theories and psychophysically testable hypotheses about higher-level visual processing. The frameworks will form the focus of discussion for workshop participants, who will include computer scientists, mathematicians and psychophysicists. Participants will evaluate, critique and discuss extensions and/or alternatives to the frameworks. The computer scientists and mathematicians in the group will provide the expertise for evaluating the sufficiency of the frameworks for formalizing and building computational theories. The psychophysicists will provide knowledge of the perceptual phenomena with which the frameworks must be tested to determine their usefulness in developing models of human vision.

Objective:

We hope to make this the first of a series of annual or semi-annual workshops focused on combining computational and psychophysical approaches to visual perception. The goal of this first workshop is to evaluate the prospects for general, formal frameworks which integrate computational and psychophysical approaches to perception. Four preliminary frameworks which have been proposed will serve as the basis for discussions in the workshop. The discussions will center on a number of critical issues revolving around the usefulness of these frameworks for the study of human vision:

• What are the computational strengths and weaknesses of the different frameworks?
• What domains of visual processing do the frameworks provide useful languages for modeling?
• What does psychophysics tell us about what is needed for a general framework of vision?
• How should the frameworks be extended or modified to increase their generality of application?
• Do the frameworks suggest new ways of asking questions for psychophysicists?
• What new experimental paradigms do the frameworks suggest for psychophysics?

Answers to these questions will lead to new ways of conceptualizing computational problems in visual perception which are amenable to psychophysical investigation.

List of participants:

Edward Adelson	Massachusetts Institute of Technology
Horace Barlow	Cambridge University
Peter Belhumeur	Harvard University
Bruce Bennett	University of California, Irvine
Michael Black	Toronto University
Andrew Blake	Oxford University
Heinrich Bulthoff	Brown University
Jacob Feldman	Rutgers University
Bill Freeman	Mitsubishi Electric Research Laboratory
Stuart Geman	Brown University
Donald Hoffman	University of California, Irvine
Allan Jepson	University of Toronto
Daniel Kersten	Universty of Minnesota
David Knill	University of Minnesota
David Kriegman	Yale University
David Mumford	Harvard University
Ken Nakayama	Harvard University
Alex Pentland	Massachusetts Institute of Technology
Ron Rensink	Harvard University
Whitman Richards	Massachusetts Institute of Technology
Philippe Schyns	Massachusetts Institute of Technology
David Sheinberg	Brown University
Shinsuke Shimojo	University of Tokyo
Michael Tarr	Yale University
John Tangney	Air Force Office of Scientific Research
Alan Yuille	Harvard University

Talk Summaries

How to Represent Data to Facilitate Association Formation

Horace Barlow
Cambridge University

It is generally agreed that the human neocortex is what makes us unique, but it is far from clear what it does for us or how it does it. Comparative anatomists have been saying since Herrick's time that it seems to give us greater knowledge of the world around us, and it can do this in two ways. First, knowledge of the world can be inherited, as numerous examples of instinctive behavior in insects and other animals proves. I don't think one should underestimate the importance of this in higher animals and humans: think of the special skills of a sheep dog, a genius like Mozart or Einstein, or the depressing lack of such skills in those under-endowed mentally. But even in these cases, what varies seems to be the capacity to acquire special or general knowledge, and this requires analysis of the noisy information provided by our senses, which is the second way referred to above.

Much of the neocortex is devoted to representing the current sensory scene. The thoughts above suggest that it may provide a representation specialized to facilitate the acquisition of knowledge. Following Shannon, the stream of sensory data can be split conceptually into information and redundancy. in spite of the pejorative term, the latter consists of all the structure, regularity and repetition in the stream of messages, and it is this part that provides us with knowledge of the world. The information by itself would be totally irregular and unpredictable and would in fact have the properties of random noise. However, the separation can only be done if everything about the structure and regularity contained in the sensory messages is known, and since knowledge is never complete, the real split is between known structure, known regularity and known repetition, and a residue that contains evidence for undiscovered, new, knowledge as well as the intrinsically unpredictable, apparently noisy information.

Three principles for a representation that would facilitate the acquisition of new knowledge will be discussed. 1) Remove evidence for the associative structures you already know about; one way of doing this is to detect conjunctions of activity in the input that occur more often than they would if randomly associated (i.e. they are "suspicious coincidences"), and use these as the representative elements passed on to the next stage. 2) Ensure that the probabilities of occurrence, as well as the actual occurrences, of the representational elements are available. 3) Ensure that the representational elements occur as far as possible independently of each other in the environment for which the representation is to be used.

Finally, I shall briefly discuss the Yellow Volkswagen problem: Can a system successfully detect associations with a conjunction of features without having elements that specifically detect that conjunction?

Lebesgue Logic and the Bayesian Foundations of Observer Theory

Bruce M. Bennett and Donald D. Hoffman
University of California, Irvine

Perceptual scientists have recently enjoyed success in constructing mathematical theories for specific perceptual capacities, capacities such as stereo-vision, auditory localization, and color perception. Analysis of these theories suggests that they all share a common mathematical structure. If this is true, the elucidation of this structure, the study of its

Our conceptual framework for this problem consists of three parts: a knowledge base, a state space of interpretations consistent with the knowledge base for the particular image, and elementary preference relations. The preference relations are used to place a partial ordering on the interpretations in the state space. This allows us to define a percept as the interpretations(s) associated with maximal nodes in the ordering.

Although conceptually simple, this definition of a percept raises many difficult issues. For example, what kinds of knowledge representations will support the reasoning process required to find consistent interpretations and maximal nodes? (We entertain one such proposal by Feldman.) Do we need to revise preference relations for each image or context? (Surprisingly, little revision may be needed.) How are priors treated? (We use only binary weights.) How are maximal nodes sought out and identified? When multiple nodes are created in the partial ordering, how then is one percept chosen over another? (Here we show how the initial percept may lead to a revision of the current ordering of the consistent interpretations.) When does it make sense to re-examine the image data and knowledge base to determine whether there is a "better" interpretation that explains more of the data? And finally, when should this process be brought to a halt? (These latter issues relate to the coherence of a percept.) For all of the above, we stress issues of competence, not performance.

Ideal observers and ideal worlds: A Bayesian view of visual information processing

David Knill and Daniel Kersten
University of Minnesota

In studying the problem of visual perception, it is necessary to decompose the general problem into small, manageable pieces. How we break up the problem and the language we use to characterize and solve sub-problems determine how well we can re-integrate these partial solutions into a general model of human perception. We argue that for problems in mid and high level vision (e.g. shape perception and object recognition), the most promising level at which to formulate problems and models is the computational level, and that the proper framework for this formulation is a Bayesian one. In this talk, we describe a particular "Bayesian" view of visual information processing based on the twin metaphors of ideal observers (a form of competence theory) and ideal worlds (a form of performance theory). The framework provides a means for a "strong" integration of computation and psychophysics in building models of perception by providing a common language for formulating models of competence and performance.

Both ideal observers and ideal worlds are characterized by posterior distributions, $p(S|I)$, specifying the probability density function for possible interpretations of a set of scene properties S, conditional on the image data I. An ideal observer does the best possible job of estimating S from I in our environment. It consists of five components: A specification of what scene properties are being estimated (perceived), a specification of the data for the estimation, a criterion for the estimation (e.g. MAP), a likelihood function (a model of image formation and image uncertainty) and a model of the prior probability density function for elements in the space of possible interpretations of S. An ideal world consists of the same five components with the exception that the likelihood function and the prior model are internalized in a human observer's visual system. An ideal world can be viewed as a description of the world in which a given human observer would be the ideal observer. Ideal observers and worlds for specific domains and problems can be incorporated into more general ideals by noting that the constituent elements of the

Instead we propose *strong coupling* in which attention is paid to the degree of dependence between the likelihood functions and prior assumptions of two sources.

Computer vision theories tend to use generic prior assumptions that are supposedly valid for a large variety of scenes. We suggest instead that it is preferable to use a set of competing specific prior assumptions geared towards the tasks that the visual system is intended to perform. We argue that this concept of *competitive priors* is desirable on theoretical grounds and is supported by experimental evidence.

Ideal observers and Psychophysics: Shape from Texture

Heinrich H. Bulthoff[†], Andrew Blake[*] and David Sheinberg[†]
†Brown University
*Oxford University

We describe an ideal observer model for estimating ``Shape from Texture" which is derived from the principles of statistical information. For a given family of surface shapes, measures of statistical information can be computed for two different texture cues - density and orientation of texels. These measures can be used to predict lower bounds on the variance of shape judgements of ``ideal" and human observers. They can also predict the optimal weights for cue integration for shape from texture. These weights are directly proportional to the information carried by each cue. The ideal observer model would thus predict that the variance of subjects' responses in a psychophysical shape adjustment task should reflect the statistical importance of individual texture cues. Our results show that human performance in shape judgements for a one-parameter family of parabolic cylinders is often better than the ideal observer using only a density cue. Therefore other information, for example the compression cue, must be used by human observers. For the first time, such results have been obtained without recourse to the unnatural cue conflict paradigms used in previous experiments. The model makes further predictions for the perception of planar slanted surfaces in the case of wide field of view.

Mid-level Vision in Scene Understanding

Edward Adelson and Alex Pentland
Massachusetts Institute of Technology

Mid level vision can use probabilistic constraints derived from the physical structure of the world in order to bridge gap between low-level primitives and higher-level scene descriptions. For instance, consider the world of "painted polyhedra," in which scenes project to images consisting of grey polygonal patches. Edges can be caused by changes in reflectance or changes in illumination (e.g., due to changes in surface normal). It is generally possible to explain a given image purely in terms of reflectance or purely in terms of illumination or by combinations thereof; a vision system must search for the "best" or "most likely" interpretation.

Local strategies, such as *junction analysis*, restrict the search but are not sufficient. We find that a global search process, involving 3-D shape recovery, junction analysis, and lighting analysis, is required to derive a stable, consistent interpretation of the scene edges and patches.

Koenderink has proposed geometric methods for determining the topologically distinct views of an object. Starting with a 3D model, this decomposition, referred to as an aspect graph, provides a complete representation of every unique view of an object. More specifically, the space of viewpoints can be partitioned into maximal regions wherein the structure of the line drawing defined by image intensity discontinuities (edges) is identical; the regions are delineated by visual events where the structure changes. The structure (topology) of the line drawing is defined by the relationship of feature points such as T-junctions, vertices, contour terminations (cusps), inflections, etc. and the smooth contours connecting them. Thus, the objects appearance is qualitatively similar for all orientations within a region; a qualitative change occurs when the orientation crosses a visual event boundary. Importantly, results from singularity and catastrophe theory indicate that there is a relatively small catalogue of visual events and consequently only a small number of ways that the image structure can change.

If humans do use multiple-views representations (even if such mechanisms are used only for particular tasks), then a principled, geometric decomposition of the view space of objects is necessary for organizing viewer-centered information. Furthermore, because representations of objects are not specified a priori in human vision, we must learn them as we explore our environment - presumably using image features similar to those specified by computational theory. Consequently, formal descriptions of object geometry, including but not limited to current aspect graph methods, offer the experimental psychologist a principled means for both manipulating the orientation of objects across surface geometry and analyzing human recognition performance and perceptual behavior.

While knowledge of the image features that define visual events is helpful in understanding object structure, it is insufficient for utilizing aspect graph models to study human shape representation. One must also have the means for decomposing actual objects into their characteristic views. This requirement has presented a major obstacle in employing such models in behavioral studies. Crucially, new results have demonstrated that Koenderink's theory is computationally tractable, and it has since enjoyed increasing popularity in the computer-vision community. Even still, the majority of work has focused on polyhedral objects; only recently have there been techniques for computing the complete aspect graphs of a variety of objects based on a combination of catastrophe theory, algebraic geometry, and robust numerical methods.

In order to assess the validity of this framework, we have initiated several studies to capitalize on these computational methods in psychophysical studies. We have begun by conducting a series of experiments to investigate whether humans are indeed sensitive to the features used in determining the topologically distinct views of an aspect graph. The subjects' task was to judge whether two consecutively presented images of the same smoothly curved object (rendered with occluding contours or shaded) were displayed at the same or at different orientations (generated by rotations in depth). Performance was assessed by measuring their accuracy in detecting an orientation difference between two images. As accuracy increases subjects are demonstrating an increased ability for discriminating a change in view. When one compares the locations of the visual events as predicted by the computational theory - that is, the orientations where the aspect graph makes the transition from one view to another - to the percent correct function, it is clear that accuracy in discriminating orientations does increase when images cross a visual event. In general, adjacent image pairs separated by visual events exhibited large increases in sensitivity.

Finally, we describe some unpublished work describing a toy world of stereoscopic figures, where only 2 or 3 disparities are defined and where a wealth of depth rankings can be discerned. Yet, each perceived configuration shows the existence of mutual constraints, not dissimilar to those originally suggested in computer algorithms to interpret scenes from line drawings.

Addendum to Chatham Bar Workshop Program

Saturday, Jan. 16

8:00 (after dinner) D. Mumford (Harvard)
  "Perception via pattern theory"

Perception via Pattern Theory

David Mumford
Harvard University

Grenander's ideas from many years ago seem to be taking on very concrete forms in recent work in computer vision and seem to be working. I would like to try to pull together his vision of the foundations of perception and contrast it with other approaches (e.g. Poggio's, Ullman's). I want to mention some extensions of his ideas: to cognitive thinking in general, to learning via minimum description length and to neural algorithms which may implement the theory.

Book Outline

Bayesian Perspectives on Visual Perception:
Computation and Psychophysics
Edited by David Knill and Whitman Richards